Bilingualism: Language and Cognition

cambridge.org/bil

Language as a pathway to promote cognitive health in older adults: effects of Between-dialect Interpreting Training (BIT) on cognitive control

Hao Wen and Yanping Dong

Language and Lifelong Development Lab, School of International Studies, Zhejiang University, Hangzhou, China

Research Article

Cite this article: Wen, H. and Dong, Y. (2025). Language as a pathway to promote cognitive health in older adults: effects of Between-dialect Interpreting Training (BIT) on cognitive control. *Bilingualism: Language and Cognition* 1–15. https://doi.org/10.1017/S1366728925100527

Received: 24 August 2024
Revised: 1 July 2025
Accepted: 26 August 2025

Keywords:
cognitive ageing; cognitive control; cognitive health; interpreter advantage; language training

Corresponding author:
Yanping Dong;
Email: ypdong@zju.edu.cn

This research article was awarded Open Data badge for transparent practices. See the Data Availability Statement for details.

Abstract

Older adults commonly experience declines in cognitive control, which significantly impacts their well-being. Although intensive language training, particularly interpreting, holds potential for mitigating these declines, its efficacy remains largely unexplored. Based on previous findings in the literature (especially our theoretical framework on interpreting), we designed a 24-hour programme of Between-Dialect Interpreting Training (BIT). Using a pretest-intervention-posttest design, we evaluated the efficacy of the BIT (over 8 weeks) against a control group on general cognitive ability (MoCA) and core cognitive control functions – working memory (via listening span and digit backward tasks), interference control (via Stroop and Flanker tasks) and cognitive flexibility (via colour-shape task and WCST). Results demonstrated notable between-group differences favouring the BIT, with significant improvements in listening span, Stroop effect and Stroop global RT, colour-shape switch cost and marginal improvements in digit backward score and MoCA. The implications of how language training promotes cognitive health during ageing are discussed.

1. Introduction

Cognitive ageing typically leads to declines in cognitive control (see Reuter-Lorenz et al., 2021 for a review), which significantly impede daily functioning for older adults (e.g., Verreckt et al., 2022) and are indicators of neurodegenerative conditions (e.g., Kirova et al., 2015). Thus, a growing body of research has focused on identifying effective strategies to promote cognitive control during ageing, particularly through cognitive training programmes that involve repeated exercises targeting specific cognitive control functions (see Wu & Rebok, 2020 for a review). However, the efficacy of these programmes, particularly in achieving far-transfer effects, remains equivocal (see Greenwood, 2020), possibly due to the lack of a robust theoretical framework grounded in training mechanisms and cognitive ageing theories (see Turnbull et al., 2022 for a unified framework for cognitive training during ageing).

One pathway to enhancing cognitive control is through language training, which represents a critical juncture in the fields of applied linguistics (e.g., Antoniou et al., 2013; Li & Dong, 2020) and cognitive enhancement (e.g., Kompa & Mueller, 2020). Language processing inherently engages a wide array of cognitive control functions (e.g., Nozari et al., 2016; Roelofs & Piai, 2011), and activates extensive brain regions, including a network of frontal, parietal and subcortical brain structures (see Ye & Zhou, 2009 for an elaborate review) that overlap with areas affected by ageing (e.g., the frontal lobe hypothesis, Reuter-Lorenz et al., 2021). The engagement of cognitive control would be intensified in more complex language activities (e.g., interpreting) due to the demanding nature of orchestrating multiple linguistic processes (see García & Kogan, 2023). The inherent coupling of language processing with cognitive control presents the potential for leveraging language training, particularly those involving intensive and complex language processing, as an effective pathway to promote cognitive health during ageing. Therefore, drawing on insights from the cognitive benefits of language training, especially from the intensive language training of interpreting, the present study leverages older adults' preserved language ability to develop an intensive language training programme that aimed at mitigating their declining cognitive control and elucidating the complex interplay between language control and cognitive control in the context of ageing.

1.1. Cognitive control in older adults

Cognitive control, also referred to as executive control or executive functioning (Diamond, 2013), involves a set of cognitive functions that are widely recognised as the fundamental driving force behind complex, goal-directed behaviour (e.g., Diamond, 2013; Miyake et al., 2000). Cognitive ageing is characterised by a decline in cognitive control, a consensus supported by major theories in the field (see Reuter-Lorenz et al., 2021 for a review). Beyond the accounts for a general decline

CAMBRIDGE UNIVERSITY PRESS

in cognitive control, recent studies suggest that this general decline is rooted in specific core deficits in cognitive control functions (e.g., Braver & West, 2008; Hasher & Zacks, 1988).

One of the most well-documented effects of ageing is observed in working memory (WM), characterised by a diminished capacity to maintain and manipulate information within the mental workspace (Braver & West, 2008; Engle & Kane, 2004). Bopp and Verhaeghen (2005) conducted a meta-analysis examining the impact of ageing on WM, showing that ageing most significantly affects complex span tasks, especially in the listening span task and the computation span task. Age-related declines also manifest in the updating function of WM, as demonstrated in tasks like the *n*-back task (Bopp & Verhaeghen, 2020) and the digit backward task (e.g., Idowu et al., 2024).

Ageing also leads to declines in interference control, marked by a reduced ability to detect and monitor conflicts (e.g., Kerns et al., 2004) and suppress irrelevant information (e.g., Hasher & Zacks, 1988), thereby resulting in greater susceptibility to distractions. Despite the theoretical assumption that ageing impairs interference control, several meta-analyses (e.g., Rey-Mermet & Gade, 2018; Verhaeghen, 2014) suggest that age-related declines in interference control tasks are mainly due to general slowing in processing speed rather than control-specific deficits. However, more recent meta-analytic evidence (Maldonado et al., 2020; Nicosia et al., 2021) challenges this view. Nicosia et al. (2021) argue that traditional methods accounting for processing speed may underestimate age-related control deficits (see also Nicosia et al., 2022, for an elaborate methodological exploration). Using multilevel modelling on the trial-level data of the Stroop task, they found that older adults did face a decline in the ability to control over prepotent interference. Similarly, Maldonado et al. (2020) used exploratory factor analyses based on the Unity-Diversity model to examine the effects of ageing on interference control. Their meta-analysis identified a significant age-related decline in the factor of interference control (Hedges' $g = 1.64$), indicating a significantly large effect of ageing on interference control processes. These findings suggest that while general slowing in processing speed does play a role, specific impairments in interference control contribute to cognitive declines in older adults, particularly in tasks requiring control over prepotent interference (e.g., Stroop task).

Alongside WM and interference control, cognitive flexibility also diminishes with age, leading to a decreased ability to switch attention between tasks and adapt to new demands and rules efficiently (see Kray & Ferdinand, 2014 for a review). This decline is evident in task-switching paradigms (see Gajewski et al., 2018 for a meta-analysis), where the mixing cost – defined as the performance difference between trials in single-task blocks and non-switch trials in mixed blocks where participants alternate between tasks – tends to increase most significantly with age. The switch cost, or the performance difference between switch and non-switch trials within mixed blocks, also increases with age but to a lesser extent. Similar reductions in cognitive flexibility are noted in rule-based switching assessed by the Wisconsin Card Sorting Test (WCST) and its variants, particularly marked by fewer completed categories and an increase in perseverative errors (see Rhodes, 2004 for a meta-analysis).

Taken together, the above analyses strongly suggest that ageing significantly impacts the cognitive control functions of WM, interference control and cognitive flexibility, making them target domains for cognitive training among older adults (e.g., Karbach & Verhaeghen, 2014).

1.2. Effects of language training on cognitive control

Language experience, particularly bilingual experience, has been assumed to confer benefits in cognitive control (e.g., Li & Dong, 2020; Bialystok & Craik, 2022). Many studies have demonstrated the cognitive benefits of managing multiple languages, particularly in domain-general cognitive control functions (see Bialystok & Craik, 2022 for a review). Lifelong experience in bilingualism has also been demonstrated to contribute to cognitive reserve, which can help maintain cognitive functions in older adults and stave off neurodegenerative conditions (see Bialystok, 2021 for an elaborate discussion). However, in recent years, the issue of bilingual advantage in cognitive control has become increasingly controversial (see Laine & Lehtonen, 2018; Paap, 2022), as reflected in the inconsistent findings from systematic reviews and meta-analyses on this topic involving both younger (see Adesope et al., 2010; Lehtonen et al., 2018) and older adults (see Degirmenci et al., 2022; Ware et al., 2020). Such inconsistencies underscore the complex and dynamic nature of bilingualism (e.g., Deluca et al., 2019), suggesting that these cognitive control benefits may be confined to specific language experiences (e.g., Paap et al., 2015; Li & Dong, 2020).

To gain a deeper understanding of the cognitive control demands associated with different language experiences, Green and Abutalebi (2013) proposed the Adaptive Control Model. This model delineates eight language control processes essential for bilingual language control: goal maintenance, conflict monitoring, interference suppression, task engagement and disengagement, selective response inhibition, salient cue detection and opportunistic planning. These processes can be effectively mapped onto domain-general cognitive control functions (Diamond, 2013; Miyake et al., 2000): maintaining and updating information (WM), suppressing competing representations (interference control) and switching between tasks (cognitive flexibility). The language control requirements for specific language experiences vary depending on three interactional contexts: the single-language context (using only one language in one environment), the dual-language context (using both languages in one environment) and dense code-switching (mixing languages within a single utterance). Notably, the dual-language context fosters the concurrent use of both languages in the same context, thereby requiring more frequent language switching (within a conversation but not an utterance, Green & Abutalebi, 2013:518). The dual-language context represents a complex linguistic environment that may engage a broader range of language control processes and places greater demands on these control processes (especially for goal maintenance and interference control), potentially requiring more diverse cognitive control phenotypes than single-language or dense code-switching contexts.

Echoing this perspective, Li and Dong (2020) further proposed the ecosystem of language experience to account for how different language learning and usage experiences may impact cognitive control. In addition to the previously mentioned interactional contexts, they suggested examining this ecosystem in terms of the intensity of language experience and its role in shaping cognitive control. Specifically, they highlighted that interpreting is a uniquely challenging bilingual task, particularly in terms of cognitive control, which goes beyond the bilingual interactional contexts outlined by Green and Abutalebi (2013). Similarly, Bialystok and Craik (2022) characterised the task of interpreting – especially simultaneous interpreting (SI) – as an intensified form of bilingualism (labelling those adept in it as super bilinguals) that

https://doi.org/10.1017/S1366728925100527 Published online by Cambridge University Press

necessitates the concurrent management of dual languages during real-time processing and is particularly taxing on the cognitive control system. Neurobiological findings further affirm the convergence of intensive language and cognitive control within interpreting. Hervais-Adelman and Babcock (2020) provided substantial evidence showing that interpreting (particularly SI) engages brain regions associated with speech perception, production, language switching, self-monitoring and decision-making. Integrating the above analyses, the language task of interpreting seems to stand out as a particularly challenging language activity that may serve as an effective training approach to enhancing cognitive control.

The connection between interpreting and cognitive control has recently been formulated in the Attention Control Model of interpreting (Dong & Li, 2020), which delineated the cognitive control challenges specific to interpreting tasks (i.e., interpreting control). According to this model, interpreting control comprises two principal components: language control and processing control. These two control processes are presumed to support interpreting tasks that require regular and frequent language switching and multitasking under time pressure. Language control is to ensure that the source language does not interfere with the target language during target language production, and this control is underpinned by cognitive control functions of WM, shifting (a simpler term for cognitive flexibility), task enhancement and disengagement (related to interference control), monitoring (related to almost all cognitive control tasks). Processing control is to ensure that the array of language tasks in interpreting are well orchestrated, and this control is supported by the cognitive control functions of coordination and WM. The Attention Control Model establishes a direct link between the core functions of cognitive control (i.e., WM, interference control, cognitive flexibility) and interpreting control (i.e., language control and processing control), corresponding precisely to the loci of cognitive declines in older adults, which offers a process-based theoretical framework guiding the development of an intensive language training programme for older adults.

1.3. Interpreter advantages in cognitive control

The focus on interpreting has led to the hypothesis that this specific language experience is especially potent in enhancing cognitive control (Dong, 2023), i.e., an interpreter advantage (for reviews, see García et al., 2020). Such an advantage is evident in core cognitive control functions, including WM (e.g., Ghiselli, 2022; Wen & Dong, 2019 for meta-analyses), interference control (see Zhao et al., 2023 for a systematic review) and cognitive flexibility (e.g., Dong, 2023, for a systematic review). It is noteworthy that most of these studies are based on young adults at their cognitive peak, an age group where bilingual cognitive control advantages are often negligible (see Ware et al., 2020). This offers promise for enhancements in cognitive control through early-stage interpreting training for older adults, who experience pronounced declines in cognitive control. While no research has specifically leveraged interpreting as an intensive language training to address cognitive ageing, empirical studies focusing on the cognitive control benefits for young trainees, particularly those at early stages of interpreting training, are highly relevant to the present study. These studies could provide valuable insights for identifying specific cognitive control challenges and potential training benefits for older adults through interpreting training.

For WM, previous research mainly focused on interpreter advantages in WM updating and WM span, which has demonstrated differential effects of an early-stage interpreting training on these two WM functions (see Dong, 2023). With a minimal amount of training (e.g., 32 academic hours plus 40 hours of after-class practice), significant enhancements in WM updating, assessed by the *n*-back task, have been observed among young trainees (mean age = 19.68, Dong et al., 2018; mean age = 19.81, Dong & Liu, 2016), and yet no advantages were found in WM span tasks (i.e., listening span and letter running span in Dong et al., 2018). Additionally, Macnamara and Conway (2014) identified similar patterns of improvement in WM among a group of 21 bimodal SI young trainees (mean age = 26.24). They observed that SI training significantly enhanced performance on the digit backward task (more related to the *n*-back task than to the complex span task; see Redick & Lindsey, 2013 for a review). However, no improvements were seen in WM span tasks (e.g., reading span and operation span). Overall, these findings suggest that the early-stage interpreting training is particularly effective in enhancing WM updating (e.g., *n*-back task, digit backward task), whereas more extensive training is needed to achieve improvements in WM span (e.g., complex span tasks).

Regarding interference control, there were only a handful of studies focusing on the interference control advantage in young interpreting trainees (see Zhao et al., 2024 for a review). Although behavioural studies have not consistently demonstrated an interpreter advantage in interference control among young trainees, often yielding null findings (e.g., Dong & Liu, 2016; Dong & Xie, 2014), two studies using event-related potential (ERP) consistently found positive evidence (Dong & Zhong, 2017; Zhao et al., 2023). Specifically, Dong and Zhong (2017) found that young trainees (mean age = 23.4) with more interpreting experience (more-IE group) demonstrated a significantly smaller flanker effect (in RTs) compared with the less-IE group (mean age = 21.2), a finding further supported by their ERP evidence. In addition, Zhao et al. (2024) also provided both behaviour (in the overall error rate) and ERP evidence for an interpreter advantage in interference control among young trainees.

The regular and frequent language switching in interpreting has also been associated with to benefits in cognitive flexibility, as typically reflected in the assessments of the colour-shape switching task and the WCST (see Dong, 2023 for a review). For the colour-shape task, an advantage in switch cost is often observed at an early stage of interpreting training, particularly in a univalent task (Dong & Liu, 2016; Zhao & Dong, 2020). However, no study has yet identified a mixing cost advantage in student trainees, and such an advantage was only observed in professional interpreters in the bivalent task (e.g., Babcock & Vallesi, 2017). As for the WCST, previous studies have consistently identified superior performance in young trainees on different indices of the WCST. Dong and Liu (2016) found an interpreter advantage in the WCST index of completed category compared with matched controls (see supplementary materials in Dong & Liu, 2016). With more interpreting training, Dong and Xie (2014) identified interpreter advantages on more WCST indices (i.e., completed category, overall error, perseverative error and previous category perseverative error) in two groups of young trainees (i.e., more-IE group: 128 hours' training; less-IE group: the triple of 128 hours' interpreting training) and the more-IE group in their study outperformed the less-IE group in the index of completed category. Taken together, interpreting training at its early stages primarily enhances trainees' performance on the WCST and the switch cost in the colour-shape task.

https://doi.org/10.1017/S1366728925100527 Published online by Cambridge University Press

To sum up, our review suggested that the early-stage interpreting training could yield significant benefits in cognitive control, with these improvements being task-specific and related to particular aspects of cognitive control. This pattern of benefits likely reflected the specific cognitive control challenges (e.g., keeping pace with the regular and frequent switching between languages, see Dong, 2023) faced by trainees during the early-stage interpreting training, and thus the cognitive control functions related to these training foci were exercised and enhanced (e.g., WM updating, local switching).

1.4. The present study

Building on the insights above, employing interpreting training appears promising for alleviating age-related cognitive control declines and thus for promoting cognitive health in older adults. However, empirical investigations into the efficacy of such training programmes are notably lacking. This absence may be due to the advanced language proficiency required for between-language interpreting, which is often impractical for older adults to achieve. Therefore, to optimise training efficacy on cognitive control and ensure its suitability for older adults, the present study is the first effort to develop an innovative language control training programme – Between-dialect Interpreting Training (BIT). While interpreting between dialects is not equivalent to between-language interpreting, the cognitive functions involved in switching between dialects engage control mechanisms analogous to those used in bilingual language control (e.g., Kirk et al., 2022). Furthermore, lifelong bidialectalism has been associated with cognitive control benefits in older adults (e.g., Hsu, 2021).

Using a pretest-intervention-posttest design, the primary objective of the present study was to assess the efficacy of the BIT programme in enhancing cognitive control among older adults, focusing on three core functions: WM, interference control, and cognitive flexibility. In addition, this study aimed to evaluate the efficacy of the BIT programme in improving older adults' general cognitive ability, an outcome widely regarded as critical for training programmes targeting this population (see Greenwood, 2020).

Regarding WM, previous research suggests that early-stage interpreting training may primarily benefit WM updating rather than WM span. Dong (2023) proposed that early-stage interpreting training may first enhance updating ability, as trainees must continuously update linguistic content to keep pace with the interpreting task (especially the CI task; see also Dong et al., 2018). Based on these findings, we predicted that the BIT programme would be effective in improving older adults' WM updating ability, whereas its impact on WM span may be more limited, at least with 24 hours of training. For interference control, although findings in the literature are somewhat mixed, two studies utilising the more sensitive ERP technique suggest that early-stage interpreting training may confer an advantage in interference control, especially in the Stroop interference task. Thus, we predicted that the BIT programme would be effective in improving older adults' interference control, with improvement potentially more pronounced in the Stroop task. As regards cognitive flexibility, previous studies have demonstrated that early-stage interpreting training effectively enhances local switching ability (indexed by switch cost) but does not significantly improve global monitoring (indexed by mixing cost). We thus predicted that the BIT programme would be effective in improving older adults' local switching control, while its impact on global monitoring control may be less pronounced. Regarding the WCST, as all previous studies using this task have reported benefits in one or more WCST indices, we predicted that the BIT programme would be effective in improving older adults' performance in the WCST. And for general cognitive ability, although no previous study has investigated the association between interpreting training and general cognitive ability, research on cognitive enhancement suggests that engaging in complex cognitive activities may contribute to overall cognitive improvements in older adults (e.g., Park & Bischof, 2013). Therefore, we predicted that the BIT programme would also be effective in enhancing older adults' general cognitive ability (measured by MoCA).

However, the majority of the above studies has focused on young adults at their cognitive peak, which stands in stark contrast to the older adults in this study, who are experiencing declines in cognitive control. Consequently, the pattern of cognitive benefits derived from an early-stage interpreting training could be influenced by this age difference. Nevertheless, the specific effects of interpreting training on cognitive control among older adults remain largely unexplored, representing a significant gap that the present study seeks to address.

2. Method

2.1. Participant

Participants in the present study were recruited through advertisements facilitated by the community office. Those interested in our programme registered at the community office and proceeded to complete a screening session, which comprised two parts: (1) A short version of the Language History Questionnaire (Li and Dong, 2020) to collect participants' background information and their language history involving both regional and standard dialects. (2) Beijing version of Montreal Cognitive Assessment (MoCA, Nasreddine et al., 2005), which was employed to evaluate participants' general cognitive status. Inclusion criteria encompassed age ($65 \leq$ age ≤ 85 years old), education level (i.e., year of education ≥ 6 years), absence of learning disabilities, no untreated severe hearing or vision impairment (self-reported) and proficient (self-reported) in both Hanzhou dialect and Mandarin Chinese. Furthermore, participants were required to meet the criterion of being free from severe cognitive impairment, determined by established cutoff scores on MoCA specifically designed for older Chinese adults. These scores took into account the participants' years of formal education, in accordance with the criteria based on Chinese elderly (Mellor et al., 2016).

Following screening, 72 participants were initially assigned to the BIT or control group (36 each) using Excel-based randomisation. Given the intensive nature of the training programme – comprising 24 formal sessions plus additional practice over a 2-month period – participants were fully informed of its demands, and voluntary participation was emphasised. As a result, two participants in the BIT group withdrew prior to training. Aligning with prior intervention protocols (e.g., Meltzer et al., 2023), these participants were replaced by two volunteers with comparable demographic and cognitive profiles to maintain group sizes. During the intervention, participants who completed less than 50% of the sessions or exhibited abnormal performance patterns were excluded from further analyses. Specifically, five participants in the BIT group withdrew due to their tight schedule in giving care to their ill spouses, difficulty following the training, or physical injury in their daily-life. In the control group, one participant was excluded due to severe emotional distress and a substantial decline in posttest performance. Accordingly,

https://doi.org/10.1017/S1366728925100527 Published online by Cambridge University Press

Table 1. Pretest group means (with SD) and comparisons of participants' background characteristics and participants' performance on each task index

Index	Control (n = 35)	BIT (n = 31)	t-value	p-value	δ_{kms}
Background characteristics					
Age	70.667 (2.828)	71.263 (3.089)	0.489	.703	0.065
Gender (F/M)	20/15	17/14	–	–	–
Education	9.167 (1.355)	10 (1.713)	1.304	.181	0.174
AOA of D2	7.024 (1.426)	7.316 (0.873)	0.609	.623	0.078
Self-rated proficiency in D1	5.571 (0.443)	5.526 (0.473)	0.249	.834	0.033
Self-rated proficiency in D2	4.667 (0.309)	4.474 (0.096)	1.529	.141	0.201
Self-rated frequency of D1 use	5.476 (0.477)	5.342 (0.473)	0.687	.496	0.013
Self-rated frequency of D2 use	3.076(0.500)	3.284 (0.459)	1.058	.297	0.089
Working memory					
Listening span score (max = 50)	19.762 (7.644)	23.789 (10.122)	1.087	.284	0.145
Digit backward score (max = 11)	4.476 (0.832)	5.316 (1.283)	1.876	.070	0.252
Interference control					
Stroop effect (RT)	140.563 (69.546)	193.155 (80.215)	1.701	.111	0.225
Stroop global RT	933.438 (147.775)	965.026 (131.272)	0.552	.584	0.072
Flanker effect (RT)	47.716 (19.495)	38.632 (18.387)	1.169	.250	0.153
Flanker global RT	553.896 (70.427)	566.079 (60.040)	0.442	.661	0.058
Cognitive flexibility					
Colour-shape switch cost (RT)	147.245 (42.855)	166.033 (45.533)	1.033	.308	0.136
Colour-shape mixing cost (RT)	108.587 (39.229)	108.267 (54.761)	0.016	.987	0.002
WCST: global ACC	.630 (0.124)	.599 (0.082)	0.697	.491	0.095
WCST: global RT	2640.687 (596.977)	2703.108 (512.568)	0.271	.788	0.004
WCST: Completed C	5.238 (1.517)	5.473 (1.235)	0.416	.680	0.054
WCST: Pers ER	14.571 (6.567)	17.158 (6.052)	0.768	.447	0.102
WCST: PC Pers ER	5.190 (3.212)	6.263 (3.564)	0.696	.491	0.131
General cognitive ability					
MoCA Score (max = 30)	25.190 (1.568)	24.211 (1.275)	1.677	.101	0.218

Notes: RT = reaction time; ACC = accuracy; C = category; Pers = perseverative; ER =error; PC = previous category; AOA = age of acquisition; D1/D2 = participants' 1st /2nd dialect; Proficiency of D1 and D2 = self-rated using a 6-point Likert Scale, with six referring to highly proficient; Frequency of use in D1 and D2: Self-rated using a 6-point Likert Scale, with six referring to highly frequent; A robust logistic regression analysis comparing gender proportions showed **no significant between-group difference** (p = .851); $\delta_{kms:}$ effect size for group differences, with the magnitude of δ_{kms} being defined as small when $0.1 \leq \delta_{kms} < 0.25$, medium when $0.25 \leq \delta_{kms} < 0.4$, and large when $\delta_{kms} \geq 0.4$.

66 participants (31 in the BIT group and 35 in the control group) were included in the final analyses. Table 1 presents their background characteristics and task performance, and Figure 1 details the recruitment, assignment, and attrition process.

2.2. Design of the Between-dialect Interpreting Training (BIT) programme

Drawing on findings from the literature, particularly the Attentional Control Model in Interpreting as our process-based theoretical framework (Dong & Li, 2020), the present study aimed to leverage the language training of interpreting as a pathway to promote cognitive health among older adults. According to the Attention Control Model, the task of interpreting exhibits two key features distinct from other language experiences: frequent and regular switching and multitasking under time pressure. These features necessitate proficient language and processing control, underpinned by core cognitive control functions. While professional-level interpreting requires specialised training, engaging in basic linguistic tasks inherent to interpreting (e.g., comprehending the source language, reformulating content across languages and producing speech in the target language) can serve as effective cognitive control training (e.g., García & Kogan, 2023).

To ensure feasibility of the training programme for older adults, the present study anchored interpreting tasks in participants' two dialects and incorporated culturally relevant materials from Hangzhou Storytelling, presented via video episodes to sustain engagement and motivation. First, the training programme was based on participants' two dialects (i.e., from Hangzhou Dialect into Mandarin Chinese), enabling our older participants to perform the cognitively demanding interpreting task and potentially enhancing training efficacy on cognitive control due to their proficiency in both dialects (see Celik et al., 2022, for the importance of proficiency in inducing cross-linguistic interference). Secondly, our training programme integrated materials from the local folk art

https://doi.org/10.1017/S1366728925100527 Published online by Cambridge University Press

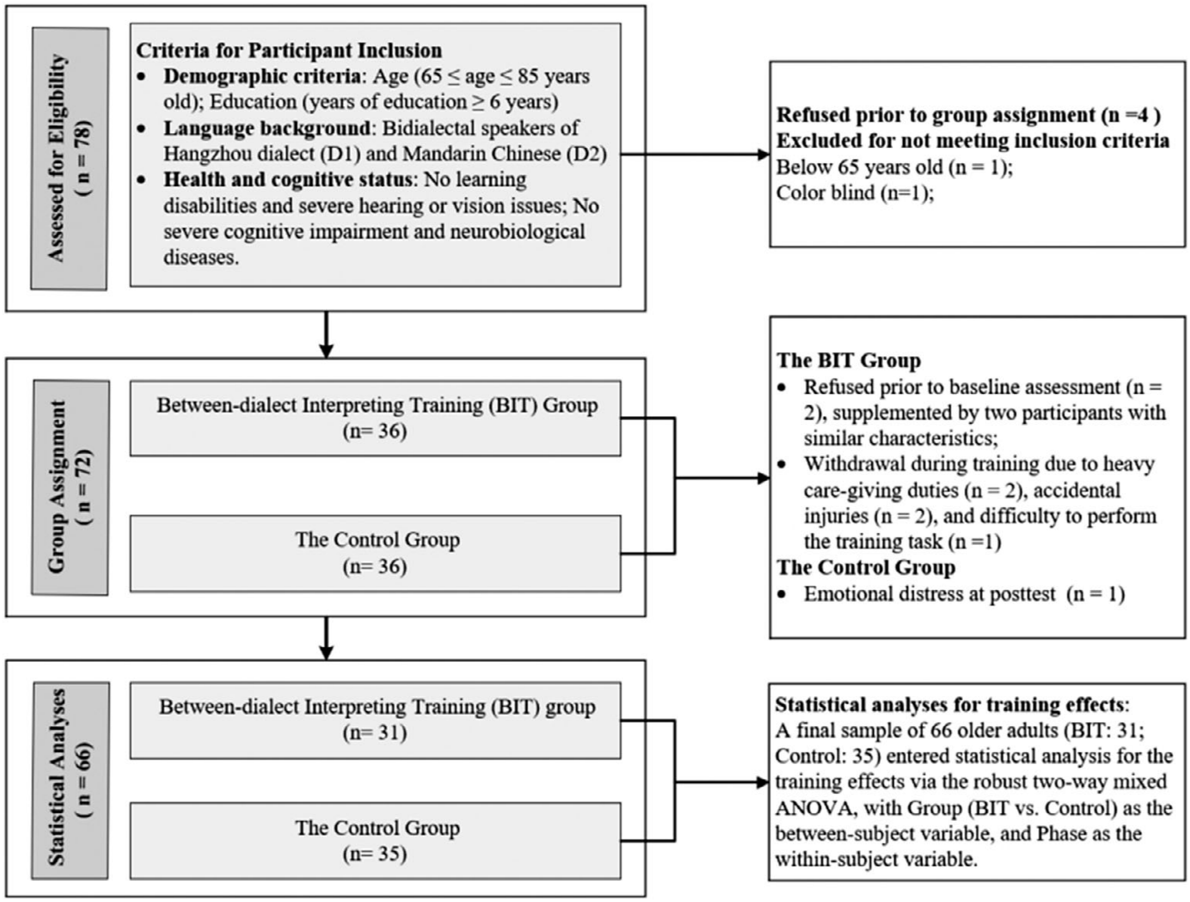

Figure 1. Process of participant recruitment, group assignment and participant attrition.

of Hangzhou Storytelling. Cultural stories from Hangzhou were specifically selected and presented to participants as video episodes performed by a Hangzhou Storytelling artist. According to findings from the pilot study, participants reported that they found the stories interesting and meaningful. Engagement with these stories helped maintain their motivation and focus on the training tasks, potentially leading to enhanced efficacy and sustained benefits over time (e.g., Wu & Rebok, 2020). Furthermore, video storytelling engaged multiple sensory stimuli (e.g., visual and auditory), fostering a rich and immersive training experience that has been suggested to be more suitable and effective for older adults (e.g., Turnbull et al., 2022).

To ensure the efficacy of the training programme, the present study specifically designed its tasks by progressively increasing task demands (moving from shadowing to SI and CI across training sessions) and by gradually introducing more complex materials (transitioning from short folk tales to longer, more complex stories from week to week, see Figure 1 for details). Participants first completed a shadowing task (in Hangzhou dialect), during which they repeated what they were listening to. What is unique is that the story did not stop until the end, and that participants had to keep pace with the story, which requires comprehension and speed. In the subsequent SI session, participants were required to "shadow" the same speech in the other dialect, which (apart from comprehension and speed) requires switching from Hangzhou dialect to Mandarin Chinese. Notably, this between-dialect SI task primarily involves phonological reformulation, with minimal demands on semantic and grammatical restructuring (see Wu et al., 2016; Wu et al., 2023), resulting in a reduced task demand compared with interlingual SI tasks. In the third session of CI, participants were required to transmit the speech segment by segment from Hangzhou dialect to Mandarin Chinese, which (apart from comprehension, speed and switching) imposes challenging demands on memory, particularly WM (see Dong et al., 2018). In addition, to better accommodate older adults' processing abilities, the training materials were carefully selected and organised by length and complexity on a weekly basis. The programme began with shorter and simpler folk tales (e.g., "Cong Bao Hui'er" in Week 1, a story about a famous snack, approximately 4 minutes) and progressed to more complex stories (e.g., "The Origin of Xiangji Temple" in Week 8, approximately 9 minutes). This approach was adopted to avoid impairing learning, which can occur if tasks are either too simple or too challenging for participants (Seitz, 2018), and to ensure an appropriate level of challenge (e.g., Lövdén et al., 2010), allowing older participants to continuously benefit from the training programme.

With the above considerations, we designed the innovative language training programme, the BIT specifically tailored for older adults. The BIT programme comprises three types of exercises: (1) shadowing in the Hangzhou dialect as a preparatory exercise to familiarize participants with the content; (2) SI from the Hangzhou dialect into Mandarin Chinese and (3) consecutive interpreting CI from the Hangzhou dialect into Mandarin Chinese. The BIT programme consists of three 1-hour sessions per week for 8 weeks.

https://doi.org/10.1017/S1366728925100527 Published online by Cambridge University Press

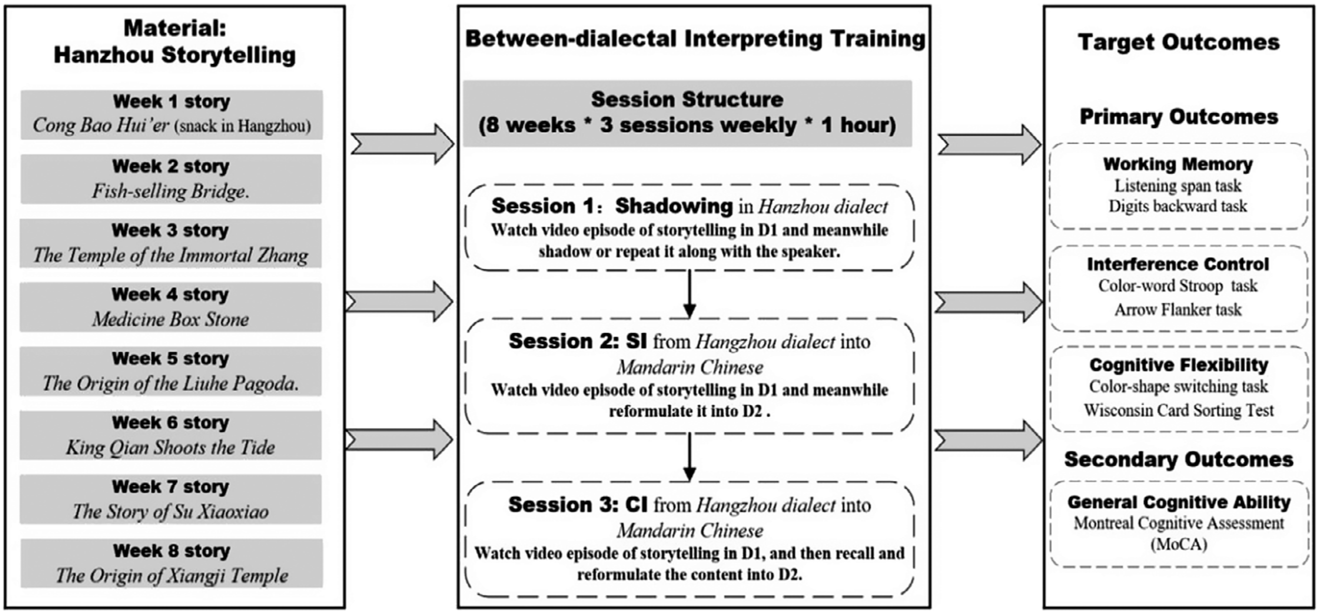

Figure 2. Material, training structure and target outcomes of the BIT programme.

This training structure adheres to principles recommended for effective cognitive control training (see Diamond & Ling, 2020 for a systematic review): spaced practice (shorter training sessions distributed over time) produces better long-term outcomes than massed practice (longer sessions condensed into a short period) and a minimum training dosage of ten sessions, each lasting at least 30 minutes, is required to achieve robust training efficacy. Figure 2 illustrates the materials, training structure and target outcomes of the BIT programme.

2.3. *Assessment*

The present study employed a battery of tasks programmed in E-Prime 3.0.1 to evaluate participants' three core cognitive control functions: WM, interference control, and cognitive flexibility. Given the complex nature of each function, we deliberately included two well-established tasks per domain to capture the nuanced effects of BIT training. These tasks are widely used in interpreting research and are known to decline with aging (as discussed in the Introduction). Additionally, participants completed the paper-and-pencil version of the MoCA to evaluate their general cognitive ability.

2.3.1. Working memory

For WM, the listening span task was employed to assess WM span (or capacity), while the digit backward span task was used to measure WM updating (e.g., Idowu et al., 2024).

The listening span task was used to assess WM span, following established paradigms used in previous studies (Unsworth et al., 2005; Dong & Cai, 2015). During the test, participants engage in a storage-plus-processing task. They first listen to a sentence and determine whether it is meaningful (i.e., the listening comprehension task), following which they hear a single digit and are required to retain it in memory (i.e., the digit recall task). Specifically, two trials (i.e., lists) were presented at each memory load level, ranging from 3 to 7 sentence-digit pairs, resulting in a total of 10 trials and a maximum possible score of 50. WM performance was measured by the total number of correctly recalled digits in serial order, with a maximum score of 50. To be included in the analysis, participants needed to achieve an accuracy level of 60%[1] on the listening comprehension task.

For WM updating, the digit backward task was applied (see WAIS; Wechsler, 2008) to assess WM updating. In this task, participants encountered digit trials ranging from 2 to 11 digits in length and were required to recall and reproduce these sequences in reverse order. At each length, two trials were presented. If both were recalled correctly, the task progressed to the next length; if only one was correct, the same length was repeated with new digit list. The task terminated when both trials at a given length were failed or when both trials at the maximum span (11 digits) were successfully completed. The digit backward score, defined as the longest length recalled, served as the measure of WM updating. The digit backward score, defined as the longest successfully recalled sequence, was used as the measure of WM updating (maximum score = 11). We chose the digit backward task over the *n*-back task, which is frequently used in interpreting studies (see Dong, 2023 for a review) due to pilot findings indicating the unsuitability of the *n*-back task for our older participants. Nevertheless, these two tasks were found to assess similar WM processes (see Redick & Lindsey, 2013 for a meta-analysis), and both are frequently used to assess updating, particularly among older adults (see Idowu et al., 2024 for a systematic review).

2.3.2. Interference control

For interference control, the colour-word Stroop task and the arrow Flanker task were selected to assess control over prepotent and distractor interference, respectively (e.g., Zhao et al., 2023).

The colour-word Stroop task is a classic measure of interference control, assessing the ability to suppress prepotent responses

[1]The present study did not follow the traditional 85% accuracy level, as such an accuracy level was questioned for excluding low-WM performers (Richmond & Burnett, 2022). Therefore, an accuracy level of 60% was selected, which, as suggested by previous studies, ensure performance above guessing probability with 95% confidence (Richmond & Burnett, 2022), and did not change the psychometric properties of the task (Đokić et al., 2018).

https://doi.org/10.1017/S1366728925100527 Published online by Cambridge University Press

(MacLeod, 1991). Participants must identify the ink colour of the colour word while ignoring its meaning. This task consists of two trial types: congruent trial (with ink colour and word matched) and incongruent (ink colour and word mismatched). Each trial followed a fixed sequence: a fixation asterisk (500 ms), a blank screen (300 ms), a stimulus display (3000 ms) and a 100 ms beep for incorrect responses, followed by a 500 ms blank screen. The congruent and incongruent trial types were presented in equal proportions (1:1) across two blocks, totalling 96 trials (48 congruent, 48 incongruent). Key indices include the Stroop effect, reflecting the difference in RTs between congruent and incongruent trials, and the Stroop global RT (i.e., the overall mean RTs across all congruent and incongruent trials) as a measure of conflict monitoring in the Stroop task.

The arrow Flanker task is a classic measure to assess the ability to resist distractors (see Dong & Zhong, 2017; Zhao et al., 2024). The task involves stimuli featuring five white arrows horizontally presented at the screen centre against a dark grey backdrop. Participants are instructed to discern whether the central target arrow points left ("<") or right (">") while disregarding non-target arrows pointing in the same or different directions as the central target (e.g., <<<<<, <<> < <). Each trial followed a fixed sequence: a fixation asterisk (350 ms), a blank screen (150 ms), a stimulus display (4000 ms) and a 100 ms beep for incorrect responses, followed by an 850 ms blank screen. Congruent and incongruent trials were presented in equal proportions (1:1) across two blocks, totalling 96 trials (48 congruent, 48 incongruent). Key indices include the Flanker effect, indicating the difference in RTs between congruent and incongruent trials and the Flanker global RT (i.e., the overall mean RTs across all congruent and incongruent trials) as a measure of conflict monitoring in the Flanker task.

2.3.3. Cognitive flexibility

For cognitive flexibility, the colour-shape switching task was used to measure cue/task-based switching, whereas the Wisconsin Card Sorting Test (WCST) assessed rule-based switching (e.g., Dong, 2023).

The univalent colour-shape switching task was adopted from the task used by Dong and Liu (2016). There were four types of stimuli in the univalent colour-shape task: two coloured patches (red/green) for colour judgments and two empty shapes (circle/triangle) for shape judgments, with the solid or dotted line serving as the cue for colour or shape judgement. Participants completed single-task blocks (judging one cue throughout) and mixed-task blocks (judging based on cues given with each stimulus). Each trial followed a fixed sequence: a fixation asterisk (350 ms), a blank screen (150 ms), a stimulus display (4000 ms), and a 100 ms beep for incorrect responses, followed by an 850 ms blank screen. The task included four single-task blocks (72 total trials) and two mixed-task blocks (96 total trials: 48 congruent, 48 incongruent), with switch and non-switch trials presented equally (1:1). We assessed two key indices: switch cost, indicating local/transient control, measured by differences in mean RTs between switch and non-switch trials in mixed blocks; and mixing cost, indicating global/sustained control, measured by differences in mean RTs between non-switch trials in mixed-task blocks and trials in single-task blocks.

The WCST was adopted from Dong and Xie (2014). The 64-item WCST version (half of the deck) was used to reduce time and difficulty for older participants based on our pilot study and previous recommendations (Purdon & Waldie, 2001). To minimise the confounding factor of pretest practice on the posttest, participants were explicitly told in the test instructions that there would be altogether three possible categorising rules (by shape, colour, or number). The WCST started with 12 practice trials, followed by 64 formal trials. For each trial, participants were required to categorize each popped-up response card (e.g., four green squares) into one of the four stimulus cards on the screen according to their shared feature in one of the three dimensions (shape, colour, and number). They were given a "correct" or "wrong" feedback for each response. The underlying categorising rule would change after a few trials (from 5 to 8 trials). At the beginning of each trial, a "+" fixation was presented for 1000 ms, after which the stimulus appeared and lasted until the participants made a response by pressing one of the four designated keys on the Chronos response box representing one of the four stimulus cards. Feedback on the accuracy of each response lasted for 500 ms, followed by the next trial until participants finished sorting all of the 64 response cards. Major WCST indices include the number of categories completed, three types of errors (i.e., overall error, perseverative error, previous category perseverative error) and global RTs.

2.3.4. General cognitive ability

The Beijing version of the MoCA was administered to evaluate general cognitive ability, following the procedure outlined by Nasreddine et al. (2005). Each participant was assigned a trained assessor who administered the MoCA in a paper-and-pencil format. This assessment consisted of various tasks evaluating cognitive domains, including visuospatial/executive functions, naming, memory, attention, language, abstraction and orientation, with a total score of 30 points. The trained assessor who administered the MoCA reviewed the participants' responses and assigned points based on the scoring criteria outlined in the MoCA manual.

2.4. Procedures

The present study employed a pretest-intervention-posttest longitudinal design to evaluate the efficacy of the BIT programme. At the pretest, participants provided informed consent and completed assessments across three sessions (around 60 minutes each). In session one, participants completed an adapted version of the LHQ for older adults and MoCA to collect their background information and general cognitive ability. During session two, participants completed three cognitive control tasks: colour-shape switching task, digit backward task and colour-word Stroop task. And in session three, participants completed another three cognitive control tasks: the listening span task, arrow Flanker task and the WCST. After the pretest, the BIT group then completed an 8-week, 24-hour training intervention (three 1-hour sessions per week) incorporating shadowing, SI, and CI exercises. Following the intervention, both BIT and control groups completed a posttest identical to the pretest.

3. Results

3.1. Data trimming

Data trimming followed standard procedures as documented in the literature. For tasks measuring RTs, data cleaning involved excluding the initial two trials per block, removing erroneous or missing responses and eliminating outliers with RTs below 200 ms or exceeding 3 SDs from the mean RTs in each participant. The total number of eliminated outliers amounted to less than 5% in each

https://doi.org/10.1017/S1366728925100527 Published online by Cambridge University Press

group, task and phase. Supplementary Material One (S1) presents details of the data trimming.

3.2. Statistical analyses

Statistical analyses were performed using R (version 4.3.2). Due to the problem of non-normal distribution and heteroscedasticity in our data, we applied robust statistical methods for all data analyses using the WRS2 package (Mair & Wilcox, 2020; Wilcox, 2023). Consistent with recommendations in robust statistical approach (e.g., Wilcox, 2017, p. 180), a default 20% trimming level was set throughout the analyses, and the statistical trimming procedures were automatically conducted through the statistical functions in the WRS2 package. This trimming level strikes a balance between ensuring sufficient statistical power for analysis (as opposed to mean, which risks low statistical power under non-normality) and mitigating the impact of extreme trimming (as opposed to 50% trimming in the median, which may discard a significant portion of data to achieve better power). This approach has been shown to perform well under non-normal conditions and with small sample sizes (e.g., Wilcox, 2022; Wilcox & Keselman, 2001).

To ensure group comparability, we applied Yuen-Welch method with a $p < .1$ threshold for group matching. With matched groups, we used a robust two-way mixed ANOVA focusing on the interaction between Group (BIT versus control) and Phase (pretest versus posttest). To further investigate specific group differences in improvements, we conducted pairwise comparisons of Phase when significant or marginal interaction effects were observed ($p < .1$). For unmatched groups ($p \le .1$), we conducted a robust analysis of covariance (ANCOVA) using pretest scores as covariates and employing trimmed means without parametric assumptions on regression line forms (e.g., Read et al., 2013). Design points were automatically selected, representing specific scores (i.e., pretest scores) where non-parametric regression lines could be compared (see Mair & Wilcox, 2020 for details). Effect sizes were calculated using the Kulinskaya et al. (2008) method (δ_{kms}). The magnitudes of δ_{kms} were categorised as small ($0.1 \le \delta_{kms} < 0.25$), medium ($0.25 \le \delta_{kms} < 0.4$) and large ($\delta_{kms} \ge 0.4$). Additionally, we have applied a Bonferroni correction to adjust significance levels for interaction effects within tasks assessing similar cognitive constructs and the significance level for pairwise comparisons when significant or marginally significant interaction effects occur. These additional analyses reached the same conclusion as obtained without the Bonferroni correction, and we have thus put it in Supplementary Material Two (S2) for the sake of better readability.

3.3. Group matching in the pretest

Before conducting the main analyses, we performed tests to assess whether the two groups were comparable in participants' background characteristics and baseline performances on all task indices. The results of these comparisons are summarised in Table 1. Specifically, no significant differences were found in participants' background characteristics (*ps* > .1), nor did participant groups significantly differ on most task indices (*ps* > .1), with the exception of a marginal difference observed in the digit backward task ($p = .07$), which led us to use the robust ANCOVA to control for the baseline difference on this specific index.

3.4. Training efficacy of the BIT programme

To clearly illustrate performance changes in descriptive data across test phases, Table 2 provides participants' performance in each task index at both pretest and posttest. As outlined in *Statistical Analyses*,

Table 2. Group means (with SD) for participants' performances in each task index at pretest and posttest

	Control (*n* = 35)		BIT (*n* = 31)	
Index	Pretest	Posttest	Pretest	Posttest
Working memory				
Listening span score (max = 50)	19.762 (7.644)	18.809 (9.597)	23.789 (10.122)	29.737(6.855)
Digit backward score (max = 11)	4.476 (0.832)	4.524 (1.222)	5.316 (1.283)	5.684 (1.230)
Interference control				
Stroop effect (RT)	140.563 (69.546)	142.645 (68.044)	193.155 (80.215)	129.352 (38.743)
Stroop global RT	933.438 (147.775)	894.480 (139.143)	965.026 (131.272)	851.153 (83.864)
Flanker effect (RT)	47.716 (19.495)	35.657 (19.980)	38.632 (18.387)	36.769 (13.642)
Flanker global RT	553.896 (70.427)	536.415 (48.975)	566.079 (60.040)	534.820 (42.923)
Cognitive flexibility				
Colour-shape switch cost (RT)	147.245 (42.855)	151.856 (44.835)	166.033 (45.533)	123.747 (43.226)
Colour-shape mixing cost (RT)	108.587 (39.229)	97.961 (30.608)	108.267 (54.761)	139.239 (43.193)
WCST: global ACC	.629 (0.130)	.686 (0.075)	.599 (0.082)	.699 (0.722)
WCST: global RT	2640.687 (596.977)	2130.235 (258.523)	2703.108 (512.568)	2148.746 (393.668)
WCST: Completed C	5.238 (1.517)	5.667 (0.901)	5.473 (1.235)	5.947 (0.912)
WCST: Pers ER	14.571 (6.567)	10.905 (4.531)	17.158 (6.052)	9.842 (4.501)
WCST: PC Pers ER	5.190 (3.212)	2.857 (2.008)	6.263 (3.564)	1.842 (1.921)
General cognitive ability				
MoCA Score (max = 30)	25.190 (1.568)	24.429 (2.356)	24.211 (1.275)	25.000 (1.250)

Notes: RT = reaction time; ACC = accuracy; C = category; Pers = perseverative; ER = error; PC = previous category.

Table 3. Summary of results for Group × Phase analyses for each task indices

	Group		Phase		Interaction effect	
	F	*p*	*F*	*p*	*F*	*p*
Working memory						
Listening span score (max = 50)	5.453	.025	2.606	.114	4.972	.031
Interference control						
Stroop effect (RT)	0.778	.383	4.056	.051	4.622	.038
Stroop global RT	0.014	.907	24.033	<.001	4.625	.021
Flanker effect (RT)	0.383	.539	3.708	.062	1.989	.167
Flanker global RT	0.065	.800	4.661	.037	0.373	.545
Cognitive flexibility						
Colour-shape switch cost (RT)	0.089	.767	4.157	.048	6.441	.015
Colour-shape mixing cost (RT)	1.911	.176	1.100	.301	4.599	.038
WCST: global ACC	0.087	.770	11.91	.001	0.929	.342
WCST: global RT	0.061	.807	31.843	<.001	0.054	.817
WCST: Completed C	0.476	.494	2.277	.139	0.006	.940
WCST: Pers ER	.174	.679	17.337	<.001	1.914	.174
WCST: PC Pers ER	.001	.976	28.951	<.001	2.766	.104
General cognitive ability						
MoCA Score (max = 30)	0.001	.710	0.141	.973	3.587	.068

Pairwise comparisons between test phases	Control (*n* = 35)	BIT (*n* = 31)
Listening span score	$M_{dif} = -0.952$, $p = .671$, $\delta_{kms} = 0.035$	$M_{dif} = 5.947$, $p = .013$, $\delta_{kms} = 0.221$
Stroop: Stroop effect	$M_{dif} = 2.081$, $p = .923$, $\delta_{kms} = 0.010$	$M_{dif} = -63.805$, $p = .009$, $\delta_{kms} = 0.325$
Stroop: Stroop global RT	$M_{dif} = -38.959$, $p = .080$, $\delta_{kms} = 0.087$	$M_{dif} = -113.873$, $p < .001$, $\delta_{kms} = 0.332$
Colour-shape: switch cost (RT)	$M_{dif} = 4.611$, $p = .717$, $\delta_{kms} = 0.034$	$M_{dif} = -42.285$, $p = .006$, $\delta_{kms} = 0.306$
Colour-shape: mixing cost (RT)	$M_{dif} = -10.627$, $p = .403$, $\delta_{kms} = 0.097$	$M_{dif} = 30.972$, $p = .052$, $\delta_{kms} = 0.202$
MoCA score	$M_{dif} = -0.762$, $p = .303$, $\delta_{kms} = 0.122$	$M_{dif} = 0.789$, $p = .057$, $\delta_{kms} = 0.201$

Notes: RT = reaction time; ACC = accuracy; C = category; Pers = perseverative; ER = error; PC = previous category; δ_{kms} = effect size for group differences, with the magnitude of δ_{kms} being defined as small when $0.1 \leq \delta_{kms} < 0.25$, medium when $0.25 \leq \delta_{kms} < 0.4$, and large when $\delta_{kms} \geq 0.4$; M_{dif} refers to the difference in trimmed mean from pretest to posttest assessment (i.e., posttest – pretest).

we applied the robust two-way mixed ANOVA to assess the training efficacy of the BIT programme on indices matched at baseline. For significant or marginal interaction effects ($p < .1$), we further conducted pairwise comparisons between test phases. Table 3 summarizes these results. For the digit backward score, given a marginal baseline difference ($p = .07$), the robust ANCOVA was used to analyse the training effect.

3.4.1. Working memory

Results of the listening span score showed significant interaction effect between Group and Phase ($p = .031$, $\delta_{kms} = 0.256$), alongside a significant main effect of Group ($p = .025$, $\delta_{kms} = 0.174$). No main effect of Phase was observed. A pairwise comparison between test phases showed significant improvement in the listening span score with the BIT group ($M_{dif} = 5.947$, $Y_t = 2.747$, $p = .013$, $\delta_{kms} = 0.221$), and such an improvement was absent in the control group ($M_{dif} = -0.952$, $Y_t = 0.430$, $p = .671$, $\delta_{kms} = 0.035$).

For the digit backward task, robust ANCOVA was employed to examine the training effect. The analysis automatically selected design points representing baseline digit backward scores (see Wilcox, 2023, p. 775 for detailed guidelines), at which posttest scores were compared (Wilcox, 1997). In our study, four design points corresponding to baseline scores of 3, 4, 5 and 6 were chosen, each meeting the minimum requirement of at least 12 samples per group at posttest (Wilcox, 1995). Figure 3 shows the ANCOVA fit for both the BIT and control groups, with the BIT group generally performing better in the posttest. Robust ANCOVA results indicated significantly better posttest performance for the BIT group at baseline score 3 ($M_{dif\ (training\ -\ control)} = 1.086$, $p = .002$, $\delta_{kms} = 0.661$), and marginal improvements at the baseline scores of 4 ($M_{dif\ (training\ -\ control)} = 0.83$, $p = .065$, $\delta_{kms} = 0.251$) and 6 ($M_{dif\ (training\ -\ control)} = 0.778$, $p = .073$, $\delta_{kms} = 0.319$). No difference was found at baseline score 5 ($M_{dif\ (training\ -\ control)} = 0.056$, $p = .891$, $\delta_{kms} = 0.021$). These results suggest that the BIT programme positively affects digit backward score, especially for participants who had low baseline scores (e.g., 3).

3.4.2. Interference control

For the Stroop task, our results demonstrated significant interaction effects in both Stroop global RT ($p = .021$, $\delta_{kms} = 0.245$) and Stroop effect ($p = .038$, $\delta_{kms} = 0.335$). There was a significant

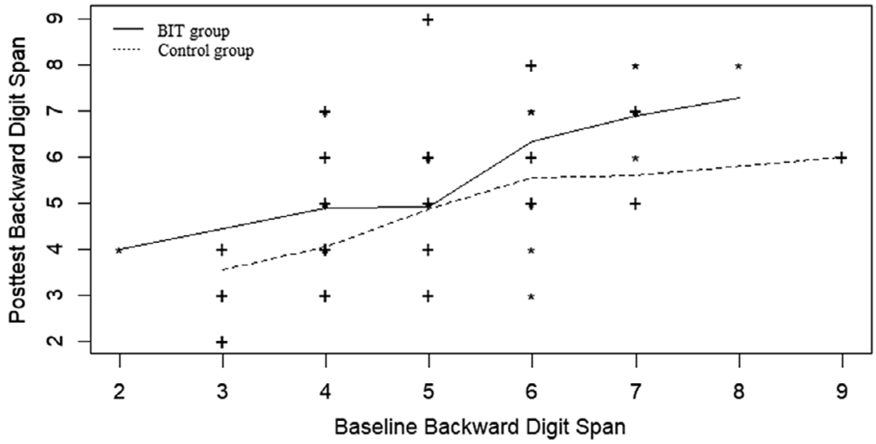

Figure 3. Robust ANCOVA fit on backward digit span score at baseline (on the *x*-axis) against posttest score (on the *y*-axis). Two regression lines represent the BIT group (solid line, plus+) and the control group (dash line, asterisk*) after smoothing.

main effect of Phase in Stroop global RT ($p < .001$, $\delta_{\text{kms}} = 0.332$) and a marginal main effect of Phase in Stroop effect ($p = .051$, $\delta_{\text{kms}} = 0.325$), with no main effect of Group in either index. Pairwise comparison of the Stroop effect indicated significant improvement in the BIT group ($M_{\text{dif}} = -63.805$, $Y_t = -2.908$ $p = .009$, $\delta_{\text{kms}} = 0.325$), with no change in the control group ($M_{\text{dif}} = 2.081$, $Y_t = 0.097$, $p = .923$, $\delta_{\text{kms}} = 0.010$). For Stroop global RT, the BIT group showed significant improvements ($M_{\text{dif}} = -113.873$, $Y_t = -4.971$, $p < .001$, $\delta_{\text{kms}} = 0.332$), while the control group only showed a marginal improvement ($M_{\text{dif}} = -38.959$, $Y_t = -1.842$ $p = .080$, $\delta_{\text{kms}} = -0.087$).

For the Flanker task, no significant interaction effect was observed neither in the Flanker effect nor in the Flanker global RTs. There was a significant main effect of Phase was observed in global RT ($p = .037$, $\delta_{\text{kms}} = 0.184$), and a marginal main effect of Phase in Flanker effect ($p = .062$, $\delta_{\text{kms}} = 0.037$). No significant main effect of Group was observed.

3.4.3. Cognitive flexibility

For the colour-shape task, we found a significant interaction effect in switch cost ($p = .015$, $\delta_{\text{kms}} = .339$). In addition, our results also identified a significant main effect of Phase in switch cost ($p = .048$, $\delta_{\text{kms}} = 0.305$), and no main effect of Group was observed. Pairwise comparison of switch cost indicated significant improvement in the BIT group ($M_{\text{dif}} = -42.285$, $Y_t = -3.115$, $p = .006$, $\delta_{\text{kms}} = 0.306$), and no significant change in the control group ($M_{\text{dif}} = 4.611$, $Y_t = 0.368$, $p = .717$, $\delta_{\text{kms}} = 0.034$).

Regarding the mixing cost, we found a significant interaction effect ($p = .038$, $\delta_{\text{kms}} = 0.299$), with no main effects of Phase or Group observed. However, contrary to our expectation, pairwise comparisons revealed that the BIT group exhibited a marginally worse performance on the index of mixing cost at posttest ($M_{\text{dif}} = 30.\ 972$, $Y_t = 2.081$, $p = .052$, $\delta_{\text{kms}} = 0.202$), while no significant difference was observed in the control group ($M_{\text{dif}} = -10.627$, $Y_t = 0.854$, $p = .403$, $\delta_{\text{kms}} = 0.097$). The increased mixing cost in the BIT group, calculated by subtracting the mean RTs in single blocks from the mean RTs in non-switch trials of mixed blocks, prompted further analyses for these two indices. Given that the Non-switch RT and Single RT are well-matched at baseline (Single RT: $p = .745$; Non-switch RT: $p = .792$), we examined participants' performance at posttest on these two indices. Specifically, a significant group difference in Single RT ($p = .048$) was noted at posttest, while no group difference was observed in Non-switch RT ($p = .845$). In addition, a significant reduction in single RT from pretest to posttest was found in the BIT group ($p < .001$), but not in the control group ($p = .151$). These results together suggest that the increased mixing cost in the BIT group is likely due to the reduced single RT, which reflects the improved general processing speed through training.

For the WCST, no significant interaction effect was observed in all indices of the WCST. However, we identified significant main effects of Phase in global accuracy ($p = .001$, $\delta_{\text{kms}} = 0.416$), global RT ($p < .001$, $\delta_{\text{kms}} = 0.384$), perseverative error ($p < .001$, $\delta_{\text{kms}} = 0.415$) and previous category perseverative error ($p < .001$, $\delta_{\text{kms}} = 0.496$). There were no significant main effects of Group in any WCST indices.

3.4.4. General cognitive ability

For general cognitive ability, we observed a marginal interaction effect in MoCA score ($p = .068$, $\delta_{\text{kms}} = 0.323$), while no main effect of Phase nor Group were observed. The pairwise comparisons further revealed a marginal improvement in the BIT group ($M_{\text{dif}} = 0.789$, $Y_t = 2.030$, $p = .057$, $\delta_{\text{kms}} = 0.201$), while no significant change was observed in the control group ($M_{\text{dif}} = -0.762$, $Y_t = -1.057$, $p = .303$, $\delta_{\text{kms}} = 0.122$).

4. Discussion

The present study is the first effort to employ interpreting training to promote cognitive health among older adults. To achieve feasibility and efficacy, we built a 24-hour BIT programme. With a pretest-intervention-posttest design for the BIT and control groups of older adults (age range: 65–81), the present study identified significant cognitive control improvements in favour of the BIT group. Specifically, the BIT group demonstrated significant improvement in WM span (indicated by an enhanced listening span score) and marginal improvement in WM updating (indicated by an enhanced digit backward score). Significant improvements were also found in interference control (evidenced by reduced Stroop effect and faster Stroop global RT) and in cognitive flexibility (reflected in reduced colour-shape switch cost). Additionally, a marginal improvement in general cognitive ability (indexed by MoCA) was observed.

4.1. Effects of the BIT on working memory

The BIT group exhibited notable improvements in WM span, as evidenced by increased listening span score, while only marginal

https://doi.org/10.1017/S1366728925100527 Published online by Cambridge University Press

improvement was observed for WM updating (indexed by digit backward task). Such a pattern, though generally supporting BIT's efficacy on WM, diverges slightly from our initial predictions (based on previous research involving younger trainees) that early-stage interpreting training would primarily enhance WM updating, with more limited effects on WM span. Most relevant to our study, Dong et al. (2018) found that early-stage interpreting training (with slightly more training than our study) did not significantly improved young trainee's listening span, although listening span did marginally predict CI performance. In addition, the training efficacy of the BIT programme on WM updating seems less robust compared with previous longitudinal studies in young trainees, which reported significant enhancements using the digit backward task (Macnamara & Conway, 2014) and the *n*-back task (Dong et al., 2018; Dong & Liu, 2016).

The differences in training efficacy between the present study and previous research could potentially be attributed to age-related cognitive differences, which may lead to varying cognitive challenges and thus different training foci. Specifically, previous research targeted young trainees at their cognitive peak, whereas our study involved older adults experiencing cognitive ageing. As mentioned in the Introduction, WM is one of the most well-documented domains of age-related decline, with complex span tasks being particularly susceptible to ageing effects (see Bopp & Verhaeghen, 2005, for a meta-analysis). Given the demanding nature of memorising and orchestrating multiple language processes during interpreting (e.g., Dong, 2023), it is possible that our older participants were especially challenged by the storage-plus-processing aspect of WM in the interpreting task and therefore prioritised overcoming these challenges to perform the training successfully. This may have led to an enhancement of their WM span. As for the marginal gains in WM updating, this could be partly explained by the adaptations made to the BIT programme for older adults. Unlike formal interpreting training for young trainees, which emphasises rapid information processing and time pressure (Dong, 2023), the BIT programme was designed to be more accessible for older adults, potentially reducing the emphasis on speeded updating processes. Consequently, it is possible that our older participants were not sufficiently challenged in this aspect of WM, leading to only marginal improvement. These age-related differences in WM challenges and in training foci during early stages of interpreting training might suggest a tentative shift in training benefits – from WM updating in younger trainees to WM span in older adults. However, this tentative shift has not yet been explicitly explored in previous research, and more empirical studies are needed to verify it.

4.2. Effects of the BIT on interference control

Consistent with our predictions regarding interference control, the present study has identified significant improvements in the Stroop task (i.e., Stroop effect and global RTs) only in the BIT group, demonstrating the efficacy of the BIT in enhancing conflict monitoring and the control of prepotent interferences. The observed benefit in conflict monitoring, indicated by smaller Stroop global RT, aligns with previous studies showing advantages for young trainees (see Zhao et al., 2023). Based on older adults, the present study extended the benefits of interpreting training to enhanced control of prepotent interference as indicated by Stroop effect, an advantage typically not observed in young trainees (see Zhao et al., 2023). Older adults face significant challenges in managing interference (Hasher & Zacks, 1988), thus making interpreting exercises that require frequent and regular language switching especially demanding for them. As analysed in Introduction, previous ERP studies attributed the lack of an interpreter advantage in interference control on behavioural measures to interpreters' superior adaptation to high-interference environments (Zhao et al., 2024), and their ability to resolve conflicts earlier in the cognitive process before overt responses (Dong & Zhong, 2017). And thus, for older adults who encounter significant challenges in managing interference (e.g., Hasher & Zacks, 1988) and experience general cognitive slowdown (Salthouse, 1991), the training effects may become more pronounced at later stages along the time course of conflict resolution, thus manifesting in behavioural measures such as the observed improvement in Stroop effect in RTs.

In contrast to the Stroop task, the present study did not identify any notable improvement in the Flanker task, neither in the Flanker effect nor in global RT. Such a contrast is likely due to different control demands in the two interference tasks, with the Stroop task being more challenging and susceptible to ageing than the Flanker task. The role of task demand has been illustrated in Xie and Dong (2017), where the intensive training of L2 public speaking produced significant interference control advantage (compared with L1 public speaking) in the more demanding task of Number Stroop, but not in the Flanker task. According to Bialystok and Craik (2022), a task that imposes more effortful demands (exceeding the control abilities of the participants, as in the Stroop task) is more likely to reveal differences between language groups compared to tasks that place relatively few demands on cognitive control (as in the Flanker task). These findings emphasise the importance of accounting for each task's unique characteristics and the cognitive demands they impose when evaluating cognitive performance in older adults.

4.3. Effects of the BIT on cognitive flexibility

For the colour-shape task, the current study observed a significant improvement of switch cost only in the BIT group, illustrating the efficacy of the BIT programme in enhancing local switching ability. As for mixing cost, although the current study observed a marginally increased mixing cost in BIT group (indicating a decline in global monitoring), our further analyses suggested that the increased mixing cost in the BIT group primarily reflects improved processing speed in single trials, rather than worsened global monitoring in non-switch trials. These findings confirmed our prediction that early-stage interpreting training is effective in enhancing local switching ability (indexed by switch cost) but not in global monitoring (indexed by mixing cost). These results also align with prior research, which has shown similar benefits on the colour-shape task during the early stages of interpreter training (Dong & Liu, 2016; Zhao & Dong, 2020). Integrating findings on interpreter advantages in cognitive flexibility, Dong (2023) hypothesised a developmental transition of cognitive flexibility from local switching control (i.e., reduced switch cost) at initial training stages (e.g., 24 hours of class training in the present study; 32 hours in Dong & Liu, 2016) to global monitoring control (i.e., reduced mixing cost) at more advanced training stages (e.g., professional levels, Babcock & Vallesi, 2017). During the initial stages of training, interpreting trainees may need to focus on overcoming challenges in alternating between listening and speaking in different languages, thereby first developing better local switching control. With more extensive training, improvements in global monitoring control (i.e., reduced mixing cost) might occur as interpreters become more adept at managing the interpreting task and have sufficient cognitive resources for global and sustained control.

For the WCST, although we expected significant gains of the BIT training, our results suggested no significant interaction effects in any of the WCST indices. As mentioned in Introduction, with slightly more training time than the present study, Dong and Liu (2016) identified an interpreter advantage specifically in the completed category. Dong and Xie (2014) also noted differences between more and less interpreting experience sub-groups only in this index. This suggests the index of completed categories is sensitive to early cognitive flexibility gains from interpreting training. However, the present study used the 64-item WCST version, shorter than the 128-item version used in the other studies, resulting in fewer categories needing sorting. This may potentially affect its sensitivity to capture improvements in the index of completed category, which has been further evidenced by a non-significant main effect of Phase (p = .139) on this index, suggesting no practice effect or training effect. In contrast, the BIT group displayed significant main effects of Phase in several other indices (see Table 3), which demonstrated a tendency towards greater gains in the posttest, as illustrated in Table 2. Despite the concern about practice effects in the WCST, the present study did hint at possible differential benefits for the BIT group (notably in global accuracy and perseverative error), with the prospect that significant interaction effects in these indices of WCST could manifest with additional training beyond the initial amount provided in our study (i.e., 24 hours' training).

4.4. Effects of the BIT on general cognitive ability

Lastly, our prediction regarding the efficacy of BIT on general cognitive ability was partially supported by the marginal improvement observed in MoCA. This may be due, in part, to the fact that the BIT programme was adapted for older adults with reduced processing demands (see Method for details), and, in part, to the paper-and-pencil format of MoCA, which may be less sensitive to detecting subtle training effects (compared with speeded measures, see Meltzer et al., 2023). Although modest, this finding suggested that sustained engagement in intensive language training can promote general cognitive benefits in older adults. Nevertheless, the marginal nature of the improvement in the present study calls for cautious interpretation, and further studies are needed to validate this finding.

4.5. Limitations and future directions

Despite the care the present study has taken in designing the BIT programme and implementing the intervention, several methodological issues need to be addressed, which could be directions for future studies.

First, the present study did not systematically monitor participants' performance progression in the trained task. Since the BIT programme primarily targets cognitive control in older adults, the training programme has been structured based on task demands (progressing from shadowing to SI and then to CI with increasing cognitive control demands, see Dong, 2023 for detailed descriptions of the cognitive demands of these tasks, and see Method for the unique demands of these tasks in the present study) and material complexity (from simpler folk tales to more complex narratives) to impose continuous cognitive challenges for our older participants. While this approach effectively improved cognitive control in our older participants, it made tracking performance progression less feasible and less critical to the current study. Nevertheless, directly assessing performance gains in the trained task would strengthen causal claims. Thus, future research with continuous assessments of the BIT task itself may further validate the efficacy of the BIT observed in the present study.

Second, the present study utilised a passive control group. While recent research advocates for active control conditions to mitigate placebo effects (e.g., Guye & Von Bastian, 2017; Nguyen et al., 2019), designing an appropriate active control for complex, multi-domain interventions like BIT are challenging. In addition, findings remain mixed on whether active controls yield significant difference with passive controls, with several recent studies yielding no meaningful difference between active and passive control (see Au et al., 2020 for a meta-analysis and a comprehensive review of this issue). Nevertheless, it is essential to systematically control for factors (e.g., placebo effects, motivation, and social interaction) that may confound interpretations of training efficacy. Doing so would strengthen the causal claims concerning the effects of BIT on targeted outcomes. Future research with well-balanced control designs is therefore expected to minimise potential interference and make causal claims with more confidence.

5. Conclusion

To sum up, the present study designed a BIT programme for older adults, and provided the first piece of evidence for how between-dialect interpreting may mitigate cognitive ageing. This BIT programme has demonstrated success in yielding domain-general cognitive benefits, extending beyond interpreting-specific skills to domain-general cognitive control functions, thereby enhancing general cognitive ability. These findings underscore the value of intensive language training as a pathway to promote cognitive health in older adults.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S1366728925100527.

Data availability statement. The data that supporting the findings of this study can be accessed via the following OSF link: https://osf.io/4dqz8/?view_only=9abb6f657ce34bc5a55c41013f5a4cde

Acknowledgements. This work was supported by a grant (22BYY077) from the Social Science Fund of China. We would like to thank the Daguan Subdistrict (Gongshu District, Hangzhou City, China) for their assistance in participant recruitment.

Competing interests. None.

Ethical standard. The present study received approval from the Ethics Committee of the School of International Studies at Zhejiang University (Review Number: SIS2022–03) and was conducted following the principles outlined in the Declaration of Helsinki, as revised in 2013.

References

Adesope, O. O., **Lavin, T.**, **Thompson, T.**, & **Ungerleider, C.** (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research*, **80**(2), 207–245.

Antoniou, M., **Gunasekera, G. M.**, & **Wong, P. C.** (2013). Foreign language training as cognitive therapy for age-related cognitive decline: A hypothesis for future research. *Neuroscience & Biobehavioral Reviews*, **37**(10), 2689–2698.

Au, J., **Gibson, B. C.**, **Bunarjo, K.**, **Buschkuehl, M.**, & **Jaeggi, S. M.** (2020). Quantifying the difference between active and passive control groups in cognitive interventions using two meta-analytical approaches. *Journal of Cognitive Enhancement*, **4**(2), 192–210.

Babcock, L., & **Vallesi, A.** (2017). Are simultaneous interpreters expert bilinguals, unique bilinguals, or both? *Bilingualism: Language and Cognition*, **20**(2), 403–417.

Bialystok, E. (2021). Bilingualism: Pathway to cognitive reserve. *Trends in Cognitive Sciences*, **25**(5), 355–364.

Bialystok, E., & **Craik, F. I.** (2022). How does bilingualism modify cognitive function? Attention to the mechanism. *Psychonomic Bulletin & Review*, **29**(4), 1246–1269.

Bopp, K. L., & **Verhaeghen, P.** (2005). Aging and verbal memory span: A meta-analysis. *Journals of Gerontology – Series B Psychological Sciences and Social Sciences*, **60**(5), 223–233.

Bopp, K. L., & **Verhaeghen, P.** (2020). Aging and *n*-back performance: A meta-analysis. *The Journals of Gerontology: Series B Psychological Sciences and Social Sciences*, **75**(2), 229–240.

Braver, T., & **West, R.** (2008). Working memory, executive control, and aging. In F. Craik & T. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 311–372). New York: Psychology Press.

Celik, S., **Kokje, E.**, **Meyer, P.**, **Frölich, L.**, & **Teichmann, B.** (2022). Does bilingualism influence neuropsychological test performance in older adults? A systematic review. *Applied Neuropsychology: Adult*, **29**(4), 855–873.

Degirmenci, M. G., **Grossmann, J. A.**, **Meyer, P.**, & **Teichmann, B.** (2022). The role of bilingualism in executive functions in healthy older adults: A systematic review. *International Journal of Bilingualism*, **26**(4), 426–449.

DeLuca, V., **Rothman, J.**, **Bialystok, E.**, & **Pliatsikas, C.** (2019). Redefining bilingualism as a spectrum of experiences that differentially affects brain structure and function. *Proceedings of the National Academy of Sciences*, **116**(15), 7565–7574.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, **64**, 135–168.

Diamond, A., & **Ling, D. S.** (2020). Review of the evidence on, and fundamental questions about, efforts to improve executive functions, including working memory. In J. M. Novick, M. F. Bunting, M. R. Dougherty, & R. W. Engle (Eds.), *Cognitive and working memory training: Perspectives from psychology, neuroscience, and human development* (pp. 143–431). Oxford University Press.

Đokić, R., **Koso-Drljević, M.**, & **Đapo, N.** (2018). Working memory span tasks: Group administration and omitting accuracy criterion do not change metric characteristics. *PLoS One*, **13**(10), e0205169.

Dong, Y. (2023). The unique bilingual profile of translators and interpreters. In A. Ferreira & J. W. Schwieter (Eds.), *The Routledge handbook of translation, interpreting and bilingualism*. New York: Routledge.

Dong, Y., & **Li, P.** (2020). Attentional control in interpreting: A model of language control and processing control. *Bilingualism: Language and Cognition*, **23**(4), 716–728.

Dong, Y., & **Cai, R.** (2015). Working memory in interpreting: A commentary on theoretical models. In Wen, Z., Mota, M. & McNeill, A. (Eds.), *Working Memory in Second Language Acquisition and Processing* (pp. 63–79). Multilingual Matters.

Dong, Y., & **Liu, Y.** (2016). Classes in translating and interpreting produce differential gains in switching and updating. *Frontiers in Psychology*, **7**, 213097.

Dong, Y., **Liu, Y.**, & **Cai, R.** (2018). How does consecutive interpreting training influence working memory: A longitudinal study of potential links between the two. *Frontiers in Psychology*, **9**(875).

Dong, Y., & **Xie, Z.** (2014). Contributions of second language proficiency and interpreting experience to cognitive control differences among young adult bilinguals. *Journal of Cognitive Psychology*, **26**(5), 506–519.

Dong, Y., & **Zhong, F.** (2017). Interpreting experience enhances early attentional processing, conflict monitoring and interference suppression along the time course of processing. *Neuropsychologia*, **95**, 193–203.

Engle, R. W., & **Kane, M. J.** (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *Psychology of Learning and Motivation*, **44**, 145–200.

Gajewski, P. D., **Ferdinand, N. K.**, **Kray, J.**, & **Falkenstein, M.** (2018). Understanding sources of adult age differences in task switching: Evidence from behavioral and ERP studies. *Neuroscience & Biobehavioral Reviews*, **92**, 255–275.

García, A. M., & **Kogan, B.** (2023). Interlingual reformulation as a window into the bilingual brain. In A. Ferreira & J. W. Schwieter (Eds.), *The Routledge handbook of translation, interpreting and bilingualism*. New York: Routledge.

García, A. M., **Muñoz, E.**, & **Kogan, B.** (2020). Taxing the bilingual mind: Effects of simultaneous interpreting experience on verbal and executive mechanisms. *Bilingualism: Language and Cognition*, **23**(4), 729–739.

Ghiselli, S. (2022). Working memory tasks in interpreting studies: A meta-analysis. *Translation, Cognition & Behavior*, **5**(1), 50–83.

Green, D. W., & **Abutalebi, J.** (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, **25**(5), 515–530.

Greenwood, P. M. (2020). Far transfer and cognitive training: Examination of two hypotheses on mechanisms. In A. K. Thomas & A. Gutchess (Eds.), *The Cambridge handbook of cognitive aging: A life course perspective* (pp. 666–684). Cambridge University Press.

Guye, S., & **von Bastian, C. C.** (2017). Working memory training in older adults: Bayesian evidence supporting the absence of transfer. *Psychology and Aging*, **32**(8), 732–746.

Hasher, L., & **Zacks, R. T.** (1988). Working memory, comprehension, and aging: A review and a new view. *Psychology of Learning and Motivation*, **22**, 193–225.

Hervais-Adelman, A., & **Babcock, L.** (2020). The neurobiology of simultaneous interpreting: Where extreme language control and cognitive control intersect. *Bilingualism: Language and Cognition*, **23**(4), 740–751.

Hsu, H. L. (2021). Cognitive control in older Minnan-mandarin and Hakka-Mandarin bidialectal adults: Advantages in Stroop-type tasks. *Lingua*, **253**, 103041.

Idowu, M. I., **Szameitat, A. J.**, & **Parton, A.** (2024). The assessment of executive function abilities in healthy and neurodegenerative aging—A selective literature review. *Frontiers in Aging Neuroscience*, **16**(1334309).

Karbach, J., & **Verhaeghen, P.** (2014). Making working memory work: A meta-analysis of executive-control and working memory training in older adults. *Psychological Science*, **25**(11), 2027–2037.

Kerns, J. G., **Cohen, J. D.**, **MacDonald, A. W.**, **Cho, R. Y.**, **Stenger, V. A.**, & **Carter, C. S.** (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science*, **303**(5660), 1023–1026.

Kirk, N. W., **Declerck, M.**, **Kemp, R. J.**, & **Kempe, V.** (2022). Language control in regional dialect speakers–monolingual by name, bilingual by nature? *Bilingualism: Language and Cognition*, **25**(3), 511–520.

Kirova, A. M., **Bays, R. B.**, & **Lagalwar, S.** (2015). Working memory and executive function decline across normal aging, mild cognitive impairment, and Alzheimer's disease. *BioMed Research International*, **2015**.

Kompa, N. A., & **Mueller, J. L.** (2020). How abstract (non-embodied) linguistic representations augment cognitive control. *Frontiers in Psychology*, **11**, 1597.

Kray, J., & **Ferdinand, N. K.** (2014). Task switching and aging. In J. A. Grange & G. Houghton (Eds.), *Task switching and cognitive control* (pp. 350–371). Oxford University Press.

Kulinskaya, E., **Morgenthaler, S.**, & **Staudte, R. G.** (2008). *Meta-analysis: A guide to calibrating and combining statistical evidence*. John Wiley & Sons.

Laine, M., & **Lehtonen, M.** (2018). Cognitive consequences of bilingualism: Where to go from here? *Language, Cognition and Neuroscience*, **33**(9), 1205–1212.

Lehtonen, M., **Soveri, A.**, **Laine, A.**, **Järvenpää, J.**, **De Bruin, A.**, & **Antfolk, J.** (2018). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological Bulletin*, **144**(4), 394–452.

Li, P., & **Dong, Y.** (2020). Language experiences and cognitive control: A dynamic perspective. *The Psychology of Learning and Motivation*, **72**, 27–52.

Lövdén, M., **Bäckman, L.**, **Lindenberger, U.**, **Schaefer, S.**, & **Schmiedek, F.** (2010). A theoretical frame-work for the study of adult cognitive plasticity. *Psychological Bulletin*, **136**, 659–676.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, **109**(2), 163–203.

Macnamara, B. N., & **Conway, A. R.** (2014). Novel evidence in support of the bilingual advantage: Influences of task demands and experience on cognitive control and working memory. *Psychonomic Bulletin & Review*, **21**, 520–525.

Mair, P., & **Wilcox, R.** (2020). Robust statistical methods in R using the WRS2 package. *Behavior Research Methods*, **52**, 464–488.

Maldonado, T., **Orr, J. M.**, **Goen, J. R. M.**, & **Bernard, J. A.** (2020). Age differences in the subcomponents of executive functioning. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, **75**(6), 31–55.

Mellor, D., **Lewis, M.**, **McCabe, M.**, **Byrne, L.**, **Wang, T.**, **Wang, J.**, **Zhu, M.**, **Cheng, Y.**, **Yang, C.**, **Dong, S.**, & **Xiao, S.** (2016). Determining appropriate

https://doi.org/10.1017/S1366728925100527 Published online by Cambridge University Press

screening tools and cut-points for cognitive impairment in an elderly Chinese sample. *Psychological Assessment*, **28**(11), 1345–1353.

Meltzer, J. A., **Kates Rose, M.**, **Le, A. Y.**, **Spencer, K. A.**, **Goldstein, L.**, **Gubanova, A.**, … **Bialystok, E.** (2023). Improvement in executive function for older adults through smartphone apps: A randomized clinical trial comparing language learning and brain training. *Aging, Neuropsychology, and Cognition*, **30**(2), 150–171.

Miyake, A., **Friedman, N. P.**, **Emerson, M. J.**, **Witzki, A. H.**, **Howerter, A.**, & **Wager, T. D.** (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, **41**(1), 49–100.

Nasreddine, Z. S., **Phillips, N. A.**, **Bédirian, V.**, **Charbonneau, S.**, **Whitehead, V.**, **Collin, I.**, **Cummings, J. L.**, & **Chertkow, H.** (2005). The Montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, **53**(4), 695–699.

Nguyen, L., **Murphy, K.**, & **Andrews, G.** (2019). Cognitive and neural plasticity in old age: A systematic review of evidence from executive functions cognitive training. *Ageing Research Reviews*, **53**, 100912.

Nicosia, J., **Cohen-Shikora, E. R.**, & **Balota, D. A.** (2021). Re-examining age differences in the Stroop effect: The importance of the trees in the forest (plot). *Psychology and Aging*, **36**(2), 214–227.

Nicosia, J., **Cohen-Shikora, E. R.**, & **Strube, M. J.** (2022). What can Brinley plots tell us about cognitive aging? Exploring simulated data and modified Brinley plots. *Frontiers in Psychology*, **13**, 1–15.

Nozari, N., **Freund, M.**, **Breining, B.**, **Rapp, B.**, & **Gordon, B.** (2016). Cognitive control during selection and repair in word production. *Language, Cognition and Neuroscience*, **31**(7), 886–903.

Paap, K. (2022). The Bilingual Advantage in Executive Functioning Hypothesis: How the Debate Provides Insight into Psychology's Replication Crisis. Routledge.

Paap, K. R., **Johnson, H. A.**, & **Sawi, O.** (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, **69**, 265–278.

Park, D. C., & **Bischof, G. N.** (2013). The aging mind: neuroplasticity in response to cognitive training. *Dialogues in Clinical Neuroscience*, **15**(1), 109–119.

Purdon, S. E., & **Waldie, B.** (2001). A short form of the Wisconsin card sorting test. *Journal of Psychiatry & Neuroscience*, **26**(3), 253–256.

Read, K. L., **Kendall, P. C.**, **Carper, M. M.**, & **Rausch, J. R.** (2013). Statistical methods for use in the analysis of randomized clinical trials utilizing a pretreatment, posttreatment, follow-up (PPF) paradigm. In J. S. Comer & P. C. Kendall (Eds.), *The Oxford handbook of research strategies for clinical psychology* (pp. 253–261). Oxford University Press.

Redick, T. S., & **Lindsey, D. R. B.** (2013). Complex span and n-back measures of working memory: A meta-analysis. *Psychonomic Bulletin & Review*, **20**(6), 1102–1113.

Reuter-Lorenz, P. A., **Festini, S. B.**, & **Jantz, T. K.** (2021). Executive functions and neurocognitive aging. In K. W. Schaie & S. L. Willis (Eds.), *Handbook of the psychology of aging* (pp. 67–81). Academic Press.

Rey-Mermet, A., & **Gade, M.** (2018). Inhibition in aging: What is preserved? What declines? A meta-analysis. *Psychonomic Bulletin & Review*, **25**, 1695–1716.

Rhodes, M. G. (2004). Age-related differences in performance on the Wisconsin card sorting test: A meta-analytic review. *Psychology and Aging*, **19**(3), 482–494.

Richmond, L. L., **Burnett, L. K.**, **Morrison, A. B.**, & **Ball, B. H.** (2022). Performance on the processing portion of complex working memory span tasks is related to working memory capacity estimates. *Behavior Research Methods*, **54**(2), 780–794.

Roelofs, A., & **Piai, V.** (2011). Attention demands of spoken word planning: A review. *Frontiers in Psychology*, **2**, 307.

Salthouse, T. A. (1991). Mediation of adult age differences in cognition by reductions in working memory and speed of processing. *Psychological Science*, **2**(3), 179–183.

Seitz, A. R. (2018). A new framework of design and continuous evaluation to improve brain training. *Journal of cognitive enhancement*, **2**(1), 78–87.

Turnbull, A., **Seitz, A.**, **Tadin, D.**, & **Lin, F. V.** (2022). Unifying framework for cognitive training interventions in brain aging. *Ageing Research Reviews*, **81**, 101724.

Unsworth, N., **Heitz, R. P.**, **Schrock, J. C.**, & **Engle, R. W.** (2005). An automated version of the operation span task. *Behavior Research Methods*, **37**(3), 498–505.

Verhaeghen, P. (2014). *The elements of cognitive aging: Meta-analyses of age-related differences in processing speed and their consequences*. Oxford University Press.

Verreckt, E., **Grimm, E.**, **Agrigoroaei, S.**, **Hubert, M.**, **Philippot, P.**, **Cremer, G.**, & **Schoevaerdts, D.** (2022). Investigating the relationship between specific executive functions and functional decline among community-dwelling older adults: Results from a prospective pilot study. *BMC Geriatrics*, **22**(1), 976.

Ware, A. T., **Kirkovski, M.**, & **Lum, J. A.** (2020). Meta-analysis reveals a bilingual advantage that is dependent on task and age. *Frontiers in Psychology*, **11**, 1458.

Wechsler, D. (2008). Wechsler Adult Intelligence Scale–Fourth Edition [dataset]. In PsycTESTS Dataset. American Psychological Association (APA). 00

Wen, H., & **Dong, Y.** (2019). How does interpreting experience enhance working memory and short-term memory: A meta-analysis. *Journal of Cognitive Psychology*, **31**(8), 769–784.

Wilcox, R. R. (1995). A regression smoother for resistant measures of location and scale. *British Journal of Mathematical and Statistical Psychology*, **48**(1), 189–204.

Wilcox, R. R. (1997). ANCOVA based on comparing a robust measure of location at empirically determined design points. *British Journal of Mathematical and Statistical Psychology*, **50**(1), 93–103.

Wilcox, R. R. (2017). *Modern statistics for the social and Behavioral sciences: A practical introduction* (2nd ed.). Taylor & Francis Group.

Wilcox, R. R. (2022). *Introduction to robust estimation and hypothesis testing*. Elsevier.

Wilcox, R. R. (2023). *A guide to robust statistical methods*. Springer Nature Switzerland.

Wilcox, R. R., & **Keselman, H. J.** (2001). Using trimmed means to compare K measures corresponding to two independent groups. *Multivariate Behavioral Research*, **36**(3), 421–444.

Wu, R., & **Rebok, G. W.** (2020). Maximizing the impact of cognitive engagement interventions for older adults. In A. K. Thomas & A. Gutchess (Eds.), *The Cambridge handbook of cognitive aging: A life course perspective* (pp. 685–700). Cambridge University Press.

Wu, Y. J., **Dang, Q.**, **Zhang, Z.**, **Zhang, H.**, & **Guo, T.** (2023). Multivariate decoding methods reveal how speaking two dialects affects executive functions. *European Journal of Neuroscience*, **57**(5), 840–853.

Wu, Y. J., **Zhang, H.**, & **Guo, T.** (2016). Does speaking two dialects in daily life affect executive functions? An event-related potential study. *PLoS One*, **11**(3), e0150492.

Xie, Z., & **Dong, Y.** (2017). Contributions of bilingualism and public speaking training to cognitive control differences among young adults. *Bilingualism: Language and Cognition*, **20**(1), 55–68.

Ye, Z., & **Zhou, X.** (2009). Executive control in language processing. *Neuroscience & Biobehavioral Reviews*, **33**(8), 1168–1177.

Zhao, H., **Chen, X.**, & **Dong, Y.** (2023). Interference control in interpreting. In A. Ferreira & J. W. Schwieter (Eds.), *The Routledge handbook of translation, interpreting and bilingualism*. New York: Routledge.

Zhao, H., **Chen, X.**, & **Dong, Y.** (2024). Young interpreting trainees' better adaptation to the flanker conflicting environment: An ERP study. *Journal of Neurolinguistics*, **70**, 101181.

Zhao, H., & **Dong, Y.** (2020). The early presence and developmental trend of interpreter advantages in cognitive flexibility: Effects from task differences and L2 proficiency. *Translation, Cognition & Behavior*, **3**(2), 241–262.

https://doi.org/10.1017/S1366728925100527 Published online by Cambridge University Press